

Final Report
Evaluating Nebraska's District Assessment Portfolios and
Recommending Model Assessments for Reading: 2000 - 2001

Prepared by

James C. Impara, Ph.D.

Chad W. Buckendahl, Ph.D.

Barbara S. Plake, Ph.D.

Buros Institute for Assessment Consultation and Outreach

A Division of the

Buros Center for Testing

University of Nebraska - Lincoln

November 2001

Final Report

Evaluating Nebraska's District Assessment Portfolios and Recommending Model Assessments for Reading: 2000 - 2001

Organization of this report

This report is in three sections. The first section describes briefly the evaluation and review process by which the district assessment portfolios were reviewed and model strategies identified. This section describes the role of the National Advisory Committee for Assessment (NACA), the criteria for evaluating district assessment portfolios, the selection and training of the District Assessment Evaluation Team (DAET), and the process for identifying model assessment strategies. The second section describes the results of the process. In the second section, the number of districts that submitted assessment portfolios is presented, as are the summary of ratings, and the results of a reliability check on the DAET. Also included in this section is a brief discussion of the model strategies.¹ The third section contains conclusions and recommendations for next year's assessment portfolio review and selection of model assessment strategies for Mathematics. There are several appendices. The appendices include the names and qualifications of the individuals who served on the District Assessment Evaluation Team (DAET) and the National Advisory Committee for Assessment (NACA), names of districts that illustrate the model assessment strategies, and data related to the reliability check on the DAET members.

Section 1: The evaluation and review process

Background

The legislation that permits Nebraska school districts to design and use their own unique assessment systems for determining the achievement levels of their students on the Nebraska content standards in reading, writing (except for certain standards assessed by the state), speaking, and listening has certain requirements. One requirement is that each district's assessment system has to be evaluated in terms of the quality of the assessments. A second requirement is that the Nebraska Department of Education (NDE) identify four models of assessments that districts may adopt or adapt in designing future assessments.

The NDE employed the Buros Institute for Assessment Consultation and Outreach (BIACO) to assist them in meeting both of these requirements. Thus, BIACO

¹ A more complete discussion is in a separate report specifically related to the models. See *Report on Model Assessment for Reading: 2000-2001*, September 2001, produced by BIACO, available from the NDE.

proposed a procedure for evaluating district assessment portfolios (descriptions of a district's assessment system) and for the identification of the four assessment models. An early decision that related to both of these activities was that the quality of district assessments would be evaluated on the extent that six technical quality criteria were met. These quality criteria are: 1) Assessments match the standards; 2) Students have an opportunity to learn assessment content; 3) Assessments are unbiased and sensitive to cultural differences; 4) Assessments are at an appropriate developmental level; 5) Assessments scores are consistent/reliable; and 6) Mastery levels are set appropriately. Each of these criteria was evaluated for each district as being Met; Met - with comment; Met - Needs Improvement; or Not met.

In April 2000, BIACO developed a draft rubric, approved by the NDE, that indicated what school districts might be expected to do to meet each criterion. Based on the overall quality of their assessment portfolio, districts are classified (rated) into one of five possible categories. These five categories are Exemplary; Very Good; Good; Acceptable, but Needs Improvement; and Unacceptable. A draft classification schedule was represented in a matrix and submitted to the NDE as part of the draft rubric in April 2000. The substance of the rubric was provided to districts in April 2001 in the form of the directions for completing their assessment portfolio. It was also displayed on the NDE web site at that time. The requirements associated with how districts would be classified (e.g., Exemplary, Very Good, etc.) were not provided to districts until summer 2001. This is because the final classification matrix was not determined until June 2001. The final matrix is shown in Appendix C.

BIACO and NDE decided that the legislative requirement for four model assessments could be interpreted as four models for each of the six technical quality criteria. Thus, the BIACO evaluation would ultimately result in 24 models. That is, there would be four models for each of the six technical quality criteria.

To accomplish the two principal outcomes of this project, evaluating district assessment portfolios and use these evaluations to identify model assessment strategies, two separate groups were formed. The first group was the National Advisory Committee for Assessment (NACA). The second group was the District Assessment Evaluation Team (DAET).

NACA

The NACA had two principal functions. The first was to provide general advice on the activities associated with the evaluation of district assessment portfolios. Second, they had to select the districts that would serve as models for each of the six technical quality criteria.

The NACA consists of four voting members and four advisory members. The four voting members were selected based on two criteria. The first criterion was that the individual be recognized nationally as having a high level of expertise in assessment, with a focus on state or local assessment programs. The second criterion was that the individual be one whose focus is on practical solutions to operational problems. These individuals were also required to be from outside Nebraska. This latter requirement reduced possible conflicts of interest when they selected the 24 models. Voting members of the NACA include a state assessment director, a local school district assessment/research director, and two university faculty members who have published on classroom assessment and have undertaken work in school systems. All members of the NACA are listed in Appendix A.

In addition to these four individuals, four other non-voting members were invited to participate. These individuals are from Nebraska. These individuals are all employees of Educational Service Units in Nebraska. The principal role of these four individuals was to provide the non-Nebraska members with an understanding about the nature of public schooling and the level of assessment literacy in Nebraska. The Nebraska members were not involved in the selection of model school districts.

The NACA met on two occasions. The first meeting was an orientation meeting in March 2001. At this meeting the Nebraska Assessment and Accountability System was described, various elements of the system that were undertaken by either BIACO or the NDE were described, and the nature and variability of the assessment literacy of Nebraska school districts was discussed. The NACA members reacted to BIACO's suggestions to the NDE regarding the technical quality rubric and the directions that were provided to school districts for completing their portfolios. These recommendations were forwarded to the NDE and considered by the NDE in their development of the district portfolio forms. The NACA also examined draft evaluation forms developed by BIACO to be used by the DAET in their evaluation of the portfolios.

The second meeting of the NACA was in August 2001. This meeting is described in detail below in the section entitled Model Selection.

DAET

A team of 16 evaluators was recruited to apply the final technical quality rubric (the rubric finally agreed upon by the NDE and BIACO in June 2000) to district assessment portfolios. This group of 16 people was named the District Assessment Evaluation Team (DAET). These individuals were recruited nationally. A major selection criterion for membership on the DAET was that the individual had a strong background in assessment (preferably a Ph.D. in assessment or a related area such as school psychology). In addition, some knowledge or experience in Nebraska was

desired, but not essential. Most DAET members had their doctorate and a strong background in assessment. Many were either from Nebraska or were familiar with Nebraska school districts (twelve were UNL graduates or had been graduate students at UNL, one was on the UNL faculty). Some had assisted districts in developing their local assessment systems. All those who assisted districts were required to evaluate assessment portfolios of districts they had not assisted. Conflicts of interest were, to the extent possible, eliminated. The names and qualifications of the DAET are shown in Appendix B.

The DAET came together in May 2000 to be trained in evaluating District Assessment Portfolios. Three districts completed drafts of their portfolios that were used in the training. The training occurred over a three-day period.

The training process included an orientation to the Nebraska assessment and accountability model and a discussion of the six quality criteria. An evaluation form was provided for the DAET members to provide feedback to districts when the process used by the district to meet a criterion needed some improvement. Each criterion was defined and discussed in an attempt to come to consensus about what the criterion meant. Once there was moderate consensus on the meaning of all the criteria, the first of the three draft district assessment portfolios was examined. In this exercise, the DAET divided into four groups of four people. Each group looked at the first quality criterion (match to standards) and, as a group completed the evaluation form. The results of each group's reactions were shared and discussed. This process was followed for each of the six quality criteria. The objective of this process was to put the initial abstract agreements about the meaning and interpretation of the criterion into an operational context. The discussion following the evaluation of each criterion often resulted in changes in what constituted acceptable procedures, what procedures might represent a need for feedback to the district, and what sorts of feedback would be appropriate to districts when feedback was needed. After the initial draft assessment portfolio was evaluated, a second was reviewed using the same process. The second evaluation further clarified the acceptable procedures associated with the criteria and the nature of appropriate feedback. Finally, a third portfolio was provided. Each DAET member examined and evaluated that individually. Following the independent evaluation, members reformed their groups and discussed their individual evaluations. These were then shared by the group, which resulted in further clarification. Because, in this first year, there were no models available for defining each of the rating criteria (a criterion was either Met; Met - with Comment; Met - Needs Improvement; or Not Met) there was some ambiguity about the interpretation of the criteria. The limited number of portfolios available for training exacerbated this problem. For this reason, two operational assessment portfolios were used to determine the extent that the DAET members were consistent in their evaluations.

During the first weeks of July 2001, after most districts had submitted their final Assessment Portfolio to the NDE, the portfolios were resent to BIACO. BIACO repackaged the Assessment Portfolios and sent them to the DAET or to a Buros Staff member for review. A few portfolios were received after the June 30 deadline that were not sent out for review, but were reviewed by Buros staff. There were a total of 19 reviewers among both the DAET and Buros staff. Most reviewers reviewed between 20 and 25² Assessment Portfolios. As a quality control measure two Assessment Portfolios were copied and sent to all reviewers. The blind (DAET reviewers did not know which districts were used for this consistency check) evaluations of these showed moderate consistency in the ratings of each of the quality criteria across all DAET and BIACO reviewers.

One element of the DAET review of each district's Assessment Portfolio was to indicate if any of the six quality criteria had been met in a particularly exemplary way. If so, that component of the Assessment Portfolio was identified as a potential model. When all districts and quality criteria were considered, over 75 districts were identified as being a potential model district for meeting at least one quality criterion. The district Assessment Portfolios that had potential models were set aside for additional review.

Model Selection

The final review to identify model strategies to meet the six quality criteria was accomplished by a National Advisory Committee for Assessment (NACA). In late August, after all reviews were completed, the NACA was convened to review the district assessment portfolios identified by the reviewers as possible models. This process took about 2 days.

One objective of the process of identifying models and districts associated with those models was to identify as many different districts as possible. Thus, there are districts that are using exemplary strategies across several of the quality criteria, but are named explicitly only once. The consequence of this was that, for some of the quality criteria it may have been possible to identify more than four models. The second consequence is that a district that is doing model work on several criteria may have been named only once.

The rating scale for each criterion was Met; Met - with comments; Met - Needs Improvement; and Not Met. Each of the models, if followed, would be classified as Met, assuming that appropriate documentation and results were provided along with the description of the procedures used to meet the criterion.

In the process of selecting models, it was noted that the descriptions in the assessment portfolios were often incomplete (most often the results of the district

² Some Buros Staff reviewed fewer than 20 Assessment Portfolios.

process were not reported), so the NACA members expanded the descriptions based on the likely results of the process. Thus, the descriptions of the models reflect district procedures in general, but the districts listed as being illustrative of that model may not be doing everything that is described for that model.

Section 2: Results of the process

This section reports the results of the evaluation of district assessment portfolios. It summarizes the ratings and describes the results of a consistency check across the portfolio evaluators. In addition, there is a summary of the results of the identification of model assessment strategies.

Evaluation of District Assessment Portfolios

Over 325 assessment portfolios were received and reviewed. Most were received by the June 30 deadline, but some arrived after that date. These portfolios represented more than 520 school districts. Fifteen consortia were submitted, as such. Although there were a number of school districts that participated in a consortium, some submitted their assessment portfolios independently. Districts that participated in a consortium submission all received the same rating. It was assumed that for most of the quality criteria (Criterion 2 – Opportunity to Learn – being the most frequent exception), all districts in a consortium used the same process and assessments.

As noted above, Assessment Portfolios for each grade level submitted were rated as being either Exemplary, Very Good, Good, Acceptable, but Needs Improvement, or Unacceptable (see the District Assessment Rating Chart in Appendix C). An Assessment Portfolio that contained descriptions of the procedures used for three grades (typically Grades 4, 8 and 11) received a separate rating for each grade. In cases where exactly the same procedures were used at each grade level, only one rating form was used. This was a frequent occurrence. In other cases, two grades (usually grades 4 and 8) used the same procedures, but the high school procedures differed. In such a case, the ratings might differ on some criteria, resulting in different overall classifications for the two grades. In yet other cases, the procedures in all three grades differed, sometimes resulting in different classifications across all three grades. Some districts did not submit Assessment Portfolios for all three grades (e.g., Class 1 districts that have only elementary grades).

An Exemplary rating was given to each Assessment Portfolio that received ratings of Met on all six quality criteria. As will be noted in the District Assessment Rating Chart in Appendix C, Met could be Met; Met - with Suggestions; or Met - Needs Improvement. The only criterion for which a distinction between Met and Met - Needs Improvement was made was Criterion 5, Score Consistency. For this Criterion, districts with a rating of less than Met, with Suggestions resulted in a lower classification. A

classification of Very Good, was given to districts that received a rating of Met on criteria one through four and a rating of less than Met - with Suggestion on criterion 5, or Not Met on Criterion 6. A rating of Good was actually difficult to obtain because most Assessment Portfolios showed sufficient quality on Criteria one through four and they Met either Criterion 5 or 6 sufficiently well to receive a rating of Very Good. Note that meeting any four criteria were not sufficient to receive a rating higher than Acceptable, but Needs Improvement. Specifically, if a district met Criteria 1 – 4, but did not meet either of Criteria 5 or 6, the rating given was Acceptable, but Needs Improvement. Many districts fell into this category because they did meet each of the first four criteria, but failed to meet both Criteria 5 and 6 (these are the two more technical criteria associated with consistency of scoring and setting mastery levels).

Table 1 shows the number of districts classified in each of the five categories. It is important to note that in Table 1 below, most of the values given for numbers of districts are approximate. This is because some districts received different ratings for their Assessment Portfolios for grades 4, 8, and 11. In other cases, a district may not have submitted portfolios for all three grades (e.g., Class 1 districts have only elementary grades).

Across all portfolios, over 90% of the districts received ratings of Acceptable, but Needs Improvement or higher. Over one-half of the districts received ratings of Very Good or Exemplary. Of the 49 districts that received ratings of Unacceptable, some received this rating because they submitted little or no documentation of the procedures and results associated with their local assessment. Others received this rating because their documentation of Criteria 1 or 2 was judged by the reviewer as being inadequate or unacceptable. The detailed ratings are shown in Table 1.

Table 1. Distribution of ratings of District Assessment Portfolios in Reading across all three grades.

Classification	Number of Districts
Exemplary	72
Very Good	228
Good	25
Acceptable, Needs Improvement	146
Unacceptable	49

As noted above, 15 assessment portfolios were received from consortia and all districts named as participants in a consortium were given the same rating because it was assumed (as per the instructions) that all districts in a consortium had used the same procedures to meet all criteria except for Criterion 2 (Opportunity to Learn). In reading the materials submitted from several of the consortia, it was clear that, although

all districts participated in various elements of the assessment activities, all districts had not followed the same procedures. This was problematic for two reasons. First, there was no provision for rating each district individually for districts submitted in a consortium, so some districts may have received a rating that was not consistent with the quality of its assessment system (its rating might have been either higher or lower). The second problem was related to the review process, in that when districts within a consortium did different things, there was no mechanism for handling that and the reviewers had to do a lot of extra work trying to sort out the most equitable rating for all districts involved.

With two exceptions, only one reviewer reviewed each Assessment Portfolio. The two exceptions were districts that were used to determine the consistency of ratings across reviewers. Because this was the first time that this process had been used, it was important to determine how consistently the reviewers rated the Assessment Portfolios. Each of the portfolios that served as the consistency check included all three grade levels. Both of these portfolios were relatively short. The portfolio labeled for purposes of this report as District A had done essentially the same thing at each grade level, suggesting that all grade levels should receive the same rating for each of the six quality criteria. District B had done different things at each of the three grade levels. The ratings for both Districts for the six quality criteria are shown in Table 2 below.

Table 2. Number of reviewers giving each rating for each criterion for two common Districts.

District	Grade Level	Criterion					
		1	2	3	4	5	6
A	All	12 Met 7 M-NI 0 NM	18 Met 1 M-NI 0 NM	18 Met 1 M-NI 0 NM	10 Met 8 M-NI 1 NM	0 Met 12 M-NI 7 NM	1 Met 7 M-NI 11 NM
B	4	0 Met 4 M-NI 15 NM	2 Met 6 M-NI 11 NM	5 Met 6 M-NI 8 NM	2 Met 2 M-NI 15 NM	0 Met 1 M-NI 18 NM	0 Met 0 M-NI 19 NM
B	8	0 Met 3 M-NI 16 NM	3 Met 5 M-NI 11 NM	5 Met 6 M-NI 8 NM	2 Met 2 M-NI 15 NM	0 Met 1 M-NI 18 NM	0 Met 0 M-NI 19 NM
B	11	0 Met 1 M-NI 18 NM	0 Met 2 M-NI 17 NM	4 Met 6 M-NI 9 NM	1 Met 2 M-NI 16 NM	0 Met 0 M-NI 19 NM	0 Met 0 M-NI 19 NM

The ratings for the six criteria for District A were the same for each of the three grade levels. However, the ratings varied in their consistency across the six quality criteria. The most consistently rated criteria for District A were Criteria 2 and 3, for which 18 of the reviewers gave the same rating (Met) and the one other reviewer gave a

rating of Met - Needs Improvement. Because the ratings of Met or Met - Needs Improvement for these two criteria did not make a difference across the final classifications, these ratings were considered highly consistent. Of more concern was the variability of ratings for Criteria 5 and 6. In the training of DAET members, there was much discussion about the requirements associated with receiving the various ratings. There were several complications for these criteria, among them were issues related to the different strategies needed when assessments were objectively scores or subjectively scores (e.g., multiple choice tests and essay tests). The techniques for the two types of tests are complex and highly technical. Clearly some reviewers interpreted the information provided by District A differently than did other reviewers. An examination of the comments provided by the reviewers suggests that the differences in perceptions of quality did not differ as much as did the rating of what was done. The comments of the reviewers for these two districts are shown in Appendix D.

Although the variability of rating for the six quality criteria was not extreme for District A, that variability resulted in substantial differences in the final classifications. This high level of variability is a direct result of the differential importance of the different criteria. Recall that to obtain a rating higher than Acceptable, but Needs Improvement, a district must be rated as being Met - Needs Improvement or higher on at least four criteria, one of which must be either Criteria 3 or 4 and the other must be either Criteria 5 or 6. Thus, meeting Criteria 1, 2, 3, and 4, but not meeting either 5 or 6 will still result in an overall classification of Acceptable, but Needs Improvement. The overall ratings for District A are shown in Table 3. These differences are due almost entirely to the Reviewers' interpretations and decisions associated with Criteria 5 and 6. Clearly, these two criteria must be better defined in the future, and substantial additional training will be needed to insure that all reviewers are interpreting them in the same way.

Table 3. Classifications for District A across Reviewers for all grades

Classification	Number of Reviewers
Exemplary	0
Very Good	8
Good	4
Acceptable, but Needs Improvement	7
Unacceptable	0

The ratings for District B are more variable for Criteria 2 and 3 and very consistent for Criteria 5 and 6. This is because there was essentially no data provided for the latter two criteria (or the District clearly misunderstood the criteria and simply provided irrelevant information). However, the critical influence of Criteria 1 and 2 as the gatekeepers for achieving any acceptable rating, resulted in much less variation in overall ratings. As shown in Table 4, across the three grades there was a high level of agreement for grades 4 and 8, and complete agreement on the final classifications for grade 11.

The relatively high consistency for District B does not preclude the need for improved training of reviewers and for high levels of agreement and understanding about what constitutes evidence of having Met the six quality criteria. Recommendations for how to improve these aspects of the project are made in the final section of this report.

Table 4. Classifications for District B across Reviewers

Classification	Number of Reviewers by grade		
	4	8	11
Exemplary	0	0	0
Very Good	0	0	0
Good	0	0	0
Acceptable, but Needs Improvement	4	3	0
Unacceptable	15	16	19

Finally, in terms of results, there are three additional areas that need to be discussed. The first relates to the late decisions about the final District Assessment Rating Chart. The Chart that was used to train reviewers differed markedly from the final Chart³. Some reviewers got the two Charts confused (they were not dated) and used the wrong chart to make their classifications. This resulted in BIACO staff having to review each Assessment Portfolio for all grades and verify that the correct classification was made.

The second area that resulted in an element of confusion was about how the district data were to be reported. Because many districts affiliated with a consortium, BIACO did not organize the data in ways that were convenient for reporting individual

³ Reviewers were provided a chart during training. They were also given the revised chart in the boxes of portfolios they received. They were also advised in a letter that the Chart had been revised.

district results. When the NDE asked for District level information, substantial effort was required to desegregate the classifications from consortia to individual districts. These procedural problems will need to be fixed in future years.

The third area is the responses by the DAET team members to a questionnaire sent to them after their ratings were completed. The questionnaire asked about the process and their recommendations for changes in the future. Their recommendations for changes next year focused mainly on the training. In summary, the DAET members who responded (12 of 16) said the training should be closer in time to the task (too much was forgotten between training and actually reviewing portfolios) and that the training needs to have better anchor (benchmark) portfolios to illustrate the different levels of quality for each criterion. When asked how much time it took to complete each portfolio review the average time was slightly more than 2 hours. Some portfolios were reviewed very quickly (those with little information and that had the same information for each grade level), however, some portfolios took up to six hours to review.

In summary, most of Nebraska's school districts submitted an assessment portfolio for review. Over 50% of the assessment portfolios obtained a rating of Exemplary or Very Good. Only 49 districts received ratings of Unacceptable. Criteria 5 and 6 were the most difficult for districts to attain because these are the most technical criteria and relatively few individuals in Nebraska have technical training in assessment.

Some confusion occurred between the NDE and BIACO in terms of what data were needed and in what format. There was also some confusion among the DAET members about what constituted the requirements to achieve various ratings.

There was some variability across the raters that may have resulted in some districts that followed essentially the same procedures being rated slightly differently. This was due, in part, to the difficulty of providing the DAET members with good illustrations for each criterion for each possible rating. Another contributing factor to the variability of ratings was that some districts that participated in a consortium did not submit their portfolios as a consortium. The individual portfolios were evaluated by different raters and may have received different ratings. DAET members also noted that they used the discussion group strategy⁴ with mixed success. Some members used it and indicated it was helpful, others tried to use it, but found it cumbersome and not helpful. They also recommended that review forms be on-line for easier completion, and that specific language (e.g., pre-coded responses) be provided for common problems to help with consistency across raters.

⁴ A moderated electronic discussion board was created to allow DAET members to ask questions about the process and get answers that all members could assess.

Summary of Model Strategies

From among the over 325 assessment portfolios submitted and evaluated, over 75 were identified as potential models for one or more of the six quality criteria. Model portfolios were developed based on the over 20 portfolios that reflected high quality assessment procedures for meeting the quality criteria. Many of the assessment portfolios were recommended as reflecting model strategies for more than one of the quality criteria. It is very possible that a district that was named as being illustrative of a model is not the only district that used the same process as described in the model. In some cases, the NACA noted that several districts (or an identified consortium of districts) used these procedures for a particular quality criterion, but even in these cases, it is likely that there are other districts that are not named that are using these same procedures.

An attempt was made to identify districts with different characteristics within each of the quality criteria. This means that, to the extent possible, small, medium, and large districts were identified as models for each quality criterion. Moreover, there was an attempt to identify four different strategies for each criterion. In some cases, the variation among the four models is only slight. For example, one district may have used a panel of local teachers to judge the match of the assessment to the standards and another district may have used a panel of local teachers supplemented by teachers from another district (used resources from outside the district).

In general, all districts that were named as illustrative of the models (see Appendix E for the names of the districts identified as using model strategies) provided complete descriptions of what they did to meet the criterion and provided results of their procedures. In many cases, the models for quality criteria 1 through 4 included some element of professional development and often used more than one procedure to meet the standard. Thus, districts used multiple methods to verify that the criterion was being met. More details of the results of the selection of model assessment portfolios is shown in the Report on Model Assessment for Reading: 2000-2001, available from the NDE.

Section 3: Conclusions and recommendations

This year was both an operational year and a learning year. As noted in the results, there were some problems that need to be overcome in the future. All of these problems can be solved.

We recommend the following as changes for the 2001-02 operational year when the assessment portfolios for mathematics will be submitted and evaluated.

1. Decisions about the classification scale and the rubric for evaluating districts should be disseminated by January 2002. This will provide the districts with a more reasonable time frame within which they can assemble their materials and organize them for the review process. The classification structure should not be changed after the DAET training has been completed.
2. School districts should be provided with additional assistance in meeting criteria 5 and 6. These were the most problematic for the districts to attain and they are the most technical. Additional workshops, web-based assistance and other resources should be used to bolster district capability in these areas. In addition, because it appeared that districts that worked closely with their ESU had higher ratings, the districts should be encouraged to employ their ESU as a resource.
3. When districts submit their materials as a consortium, that each district in the consortium should use the same procedures for all relevant criteria (i.e., all criteria except Criterion 2 – Opportunity to Learn). If this is not the case, then each district should submit their assessment portfolios independently.
4. DAET training should be modified in several ways. First, it should occur about the same time as the ratings are to be made. Second, marker (benchmark) portfolios should be provided to illustrate the different ratings (e.g., what constitutes a Met for criterion 5). Third, the communications strategy for the DAET should be improved. Fourth, all ratings should be submitted electronically, preferably on-line. Finally, a set of frequently expected responses should be developed and provided to DAET members to use when certain problems are encountered.
5. DAET members should receive appropriate payment for the time spent. The amount paid this year was low considering the amount of work done. In future years, as districts become more capable, the quality of the portfolios is expected to increase along with the amount of information provided in the portfolio.
6. Improved data record management files should be set up so that the NDE can obtain the data it needs to provide feedback and rating information to districts and so that summary data can be more readily retrieved.
7. There are many “small” things that need to be modified and adjusted as the program is continued (e.g., methods for transporting and storing the assessment portfolios).

Overall we believe the process worked well in this first operational year. Many decisions were made late in a process that was new to everyone involved. As the Nebraska Assessment and Accountability System continues to evolve, it is expected to operate much more smoothly and efficiently.

Appendix A

Names and qualifications of National Advisory Committee for Assessment

Name	Current Position	Other qualifications
Dr. Jeri Benson	Professor, University of GA. (GA)	Editor of major journal in applied measurement, expert in test validity.
Dr. Susan Brookhart	Professor, Duquesne University (PA)	Has major publications related to classroom assessment
Dr. Doug Rindone	Chief, Bureau of Evaluation & Student Assessment CT Dept Of Education (CT)	Works with several other states on their assessment advisory committees
Dr. Joe Wilhoft	Director, Planning, Assessment, and Evaluation, Tacoma Public Schools (WA)	Works with State of Washington Assessment (Technical Advisory Committee, and other school districts.
Penny K. Businga	ESU 13 (NE)	Works with districts in the ESU 13 service unit on assessment tasks.
Dr. Donald L. Fritz	ESU 6 (NE)	Works with districts in the ESU 6 service unit on assessment tasks.
Dr. Leslie Lukin	ESU 18, Lincoln Public Schools (NE)	Conducts workshops statewide on assessment topics.
Carla J. Noerrlinger	ESU 19, Omaha Public Schools (NE)	Supervised development of OPS Benchmark assessment program

Appendix B

Names and qualifications of District Assessment Evaluation Team

Name	Current Position	Other qualifications
Dr. Debbi Bandalos	Associate Professor, UN-L (NE)	Works with NE school districts on assessment development
Dr. Laura Barnes	Associate Professor, OK. State University (OK)	UN-L Graduate, teaches measurement courses, works with test publishers on norming projects
Dr. Jennifer Fager	Director, Curriculum and Assessment, Central Michigan University (MI)	UN-L Graduate, has worked on assessment related projects with schools.
Dr. Julie Flegel	Psychometrician, US Postal Service (OK)	Taught courses in measurement, worked with teacher testing program
Dr. Kathleen Itzen	Private Consultant (NE)	UN-L Graduate, worked with districts on assessment development, participated in alignment studies.
Dr. Gerald Giraud	Assistant Professor, Nebraska Methodist University (NE)	UN-L Graduate, has worked on measurement projects with NE school districts.
Jeanette Jackson	Retired (NE)	Previously worked at ESU 7, evaluated District Assessment Plans for NDE.
Dr. Jessica Jonson	University Wide Assessment Coordinator, UN-L (NE)	UN-L Graduate, works across all UN-L departments to assess student outcomes.
Dr. Sherral Miller	Program Director, ACT (IA)	UN-L Graduate, has worked on many assessment projects with schools and other agencies.

Dr. Lori Nebelsick-Gullett	Director, Testing and Evaluation, Richardson (TX) Public Schools (TX)	UN-L Graduate. Ten years experience in working with state and local testing programs.
Name	Current Position	Other qualifications
Dr. Heidi Paa	Assessment Developer, Renaissance Learning, Inc. Formerly, Assistant Professor, University of Wisconsin (WI)	UN-L Graduate, worked as assessment coordinator at UW-Madison.
Dr. June Smith	Manager, Research Services, National Council of State Boards of Nursing (IL)	UN-L Graduate, develops assessments for nurse licensure and classroom tests.
Dr. Howard Stoker	Retired (TN)	University faculty for over 30 years, directed a state testing program in FL, has worked with numerous state, and local, agencies on testing issues.
Dr. Elisabeth Sundermeier	Staff Counselor, Rhode Island Counseling Center (RI)	UN-L Graduate, worked in Buros fact checking test reviews.
Dr. James Swanson	Retired (CA)	Assisted in design of State Assessment program (FL), taught measurement, developed items for state and local assessment programs.
Dr. Ed Wiley	Project director, American Institutes for Research (CA)	UN-L Graduate (MA), worked on a variety of assessment related project contracts.

DISTRICT ASSESSMENT RATING CHART

2000-2001

Assessment Rating	Exemplary	Very Good	Good	Acceptable, But Needs Improvement	Unacceptable
1. Matches Standards	Met	Met	Met	Met	Met ⁵ or Met
2. Aligns with Curriculum	Met	Met	Met	Met	Met
3. Bias Review	Met	Met	Met	Not Met	Not Met
4. Developmentally Appropriate	Met	Met	Met	Not Met	Not Met
5. Score Consistency	Met	Met	Met - Needs Improvement	Not Met	Not Met
6. Mastery Levels	Met	Met	Met	Not Met	Not Met
Met could be "Met", "Met - with Suggestions", or "Met - Needs Improvement"					

⁵ An arrow indicates that the either of the two criteria joined by the arrow must be met.

Appendix D

Summarized Comments from Reviewers for each Criterion for Each of Two Districts Rated by All Reviewers and Number of Reviewers Making Each Comment

District A

All grade levels were the same and were combined by all reviewers.

Criterion 1: Alignment to Standards

- 10 - Description of the process and results is needed.
- 6 - Provide evidence of sufficient coverage.
- 2 - Suggested a group independent of the development team review alignment.
- 1 - Included all language from Appendix B.
- 6 - No comments.

Criterion 2: Opportunity to Learn

- 6 - Requested additional documentation on panel and process.
- 1 - Insufficient evidence for all choices that were marked on the portfolio.
- 12 - No comments.

Criterion 3: Freedom from Bias/Sensitive Language

- 6 - More information about the process used to conduct bias review.
- 3 - Present the results of bias review and any decisions.
- 1 - Suggested consulting the assessment's technical manuals.
- 9 - No comments.

Criterion 4: Developmentally Appropriate Level

- 10 - More information about the panel and process used.
- 2 - Suggested that the results of the review and any decisions be presented.
- 1 - Suggestion that readability and task difficulty be considered because they were not included in the portfolio.
- 6 - No comments.

Criterion 5: Consistency in Scoring

- 12 - Recommended that results be presented.
- 8 - Needed to clarify what two assessments were compared.
- 5 - Suggested strategies for both objectively and subjectively scored assessments.
- 4 - Clarify if scoring rubric had been pre-tested and agreement estimated.
- 2 - Indicated that there was a lack of documentation of the process.
- 2 - Suggested that they confused the information with criterion 6.
- 1 - Suggested additional literature.
- 1 - Included all language from Appendix B.

Criterion 6: Mastery Levels are Appropriate

- 17 - Description of process needs evidence that difficulty was considered.
- 4 - Description of the rubric method needed.
- 2 - Suggested strategies for both objectively and subjectively scored assessments.
- 1 - Expand the description of the panel's qualifications.
- 1 - Included all language from Appendix B.
- 1 - No comments.

District B

Grade levels differed, but comments are combined across grade levels.

Criterion 1: Alignment to Standards

- 13 - Description of the process and results is needed or is unclear.
- 7 - Provide evidence of sufficient coverage.
- 4 - Suggested a group independent of the development team review alignment.
- 3 - No documentation provided.
- 1 - Included all language from Appendix B.
- 1 - Unclear as to what the number five written by number of multiple choice items refers to.
- 4 - No comments.

Criterion 2: Opportunity to Learn

- 8 - Requested additional documentation on panel and process.
- 4 - Results needed.
- 6 - Lacking description of process used.
- 1 - Insufficient evidence for all choices that were marked on the portfolio.
- 1 - Suggested using a matrix or survey of when skills were taught and tested.

- 1 - No documentation provided about qualifications of panel.
- 3 - No comments.

Criterion 3: Freedom from Bias/Sensitive Language

- 9 - More information about the process used to conduct bias review.
- 4 - Present the results of bias review and any decisions.
- 4 - Questioned use of Stiggins' book on *Student Centered Classroom Assessments* to address bias/offensive situations – need more description.
- 2 - Suggested next time get a process in place and an appropriate panel.
- 2 - Description of the qualifications of the trainers needed.
- 1 - Included all language from Appendix B.
- 1 - Description of teacher training needed.
- 5 - No comments.

Criterion 4: Developmentally Appropriate Level

- 6 - More information about the panel and process used.
- 6 - Lacking description of process used.
- 4 - Suggested that the results of the review and any decisions be presented.
- 2 - Process inadequate (one teacher determine and judge assessments), suggested using a panel of independent reviewers.
- 1 - Suggested looking at quality criterion checklist.
- 1 - Provide information about action taken to correct assessments not at an appropriate level.
- 4 - No comments.

Criterion 5: Consistency in Scoring

- 14 - Indicated that there was a lack of documentation of the process.
- 8 - Recommended that results be presented.
- 4 - Suggested strategies for both objectively and subjectively scored assessments.
- 3 - Description of staff training process needed.
- 2 - Inappropriate methods used to estimate consistency.
- 1 - Suggested getting expert advice on how to demonstrate reliability in a psychometric context.
- 1 - Need clarification on double scoring and percent of scorer agreement.
- 1 - Included all language from Appendix B.
- 4 - No comments.

Criterion 6: Mastery Levels are Appropriate

- 10 - No documentation or information provided.
- 5 - Description of process needs evidence that difficulty or students' characteristics were considered.
- 2 - Administrator's judgment not sufficient.
- 1 - Suggested getting expert advice in either establishing a process or describing the existing process.
- 1 - Included all language from Appendix B.
- 1 - Neither of the recommended methods were used.
- 1 - Expand the description of the panel's qualifications.
- 5 - No comments.

Appendix E

Criterion	District(s) that are illustrative of the model
1	Bellevue
1	Beatrice & Omaha
1	Hampton
1	Waterloo
2	Elkhorn
2	Crawford
2	Ralston
2	Hanover & Nemaha Valley
3	Valley, Blair, Bennington, Conestoga, Johnson-Brock
3	Lincoln
3	Winnebago
3	David City
4	Bancroft-Rosalie, Mead, Scribner-Snyder
4	Crete, Ashland-Greenwood
4	Hastings
4	McCool Junction
5	Niobrara
5	Millard
5	Cheney
5	Raymond Central, Yutan
6	Ralston
6	Panhandle consortium (ESU 13)
6	Platteville, North Bend Central
6	Medicine Valley